



Strathprints Institutional Repository

Suberu, John and Gromski, Piotr S. and Nordon, Alison and Lapkin, Alexei (2016) Multivariate data analysis and metabolic profiling of artemisinin and related compounds in high yielding varieties of *Artemisia annua* field-grown in Madagascar. Journal of Pharmaceutical and Biomedical Analysis, 117. pp. 522-531. ISSN 0731-7085 , <http://dx.doi.org/10.1016/j.jpba.2015.10.003>

This version is available at <http://strathprints.strath.ac.uk/54701/>

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: strathprints@strath.ac.uk



Multivariate data analysis and metabolic profiling of artemisinin and related compounds in high yielding varieties of *Artemisia annua* field-grown in Madagascar

John Suberu^a, Piotr S. Gromski^b, Alison Nordon^b, Alexei Lapkin^{a,*}

^a Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB2 3RA, UK

^b WestChem, Department of Pure and Applied Chemistry, and Centre for Process Analytics and Control Technology (CPACT), University of Strathclyde, Glasgow G1 1XL, UK

ARTICLE INFO

Article history:

Received 10 May 2015

Accepted 1 October 2015

Keywords:

Artemisia annua

Multivariate data analysis

Metabolic profiling

ABSTRACT

An improved liquid chromatography–tandem mass spectrometry (LC–MS/MS) protocol for rapid analysis of co-metabolites of *A. annua* in raw extracts was developed and extensively characterized. The new method was used to analyse metabolic profiles of 13 varieties of *A. annua* from an in-field growth programme in Madagascar. Several multivariate data analysis techniques consistently show the association of artemisinin with dihydroartemisinic acid. These data support the hypothesis of dihydroartemisinic acid being the late stage precursor to artemisinin in its biosynthetic pathway.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In 2013, there were an estimated 198 million cases of malaria globally, resulting in over half a million fatalities [1]. These numbers are a significant improvement on the morbidity and mortality data from previous years [2–4] due, in part, to the effectiveness of the artemisinin combination therapies (ACTs), backed by the World Health Organisation (WHO). ACT is used in the treatment of uncomplicated *P. falciparum* malaria. The active ingredients of ACTs are derived from artemisinin (1), an endo-peroxide lactone obtained from *Artemisia annua* (*A. annua*) plant. Artemisinin is now also being produced semi-synthetically to improve the stability of supply of this active pharmaceutical ingredient (API) [5].

However, the natural product route to artemisinin remains an important part of the supply chain. To keep the plant route profitable necessitates the development of high-yielding plant cultivars through selective breeding programmes. If biorefining will in the future develop into industrial reality, it will offer additional opportunities to increase the value of plant production, e.g. through co-production of other valuable secondary metabolites that could be converted to artemisinin, or used in the manufacture of other functional molecules of commercial interest [6]. Identifying potential target compound(s) in plant extracts will then need to take

into account the biosynthesis of artemisinin, the concentration of metabolites and the mechanism by which these precursors are converted to artemisinin *in planta*.

Biosynthesis of artemisinin has been well studied and several reviews on the topic are published [7–10]. Similar to all sesquiterpenes, the synthesis involves condensation of 5-carbon molecules from the mevalonate and the deoxy-D-xylulose-5-phosphate (DXP) pathways to farnesyl pyrophosphate (IPP) as shown in Fig. 1. In *A. annua* the mevalonate and the DXP pathways occur in cytosol and plasmid, respectively. The committed step of the biosynthesis is the cyclization of farnesyl diphosphate into amorpha-4,11-diene and its subsequent conversion to artemisinin alcohol. The route from artemisinin alcohol to artemisinin is still not entirely clear: published data is inconclusive and sometimes contradictory [11,12]. The final stages of artemisinin biosynthesis remain controversial with conflicting theories identifying the last-stage precursor either as dihydroartemisinic acid (4) or artemisinic acid (5). Li et al. [11] has reviewed evidence for artemisinic acid, while Brown and others [7,9,13,14] have done the same for dihydroartemisinic acid as the key precursors Scheme 1.

In our previous studies we have investigated the variability of metabolic profiles of *A. annua* grown in very different geographical regions and the impact of this variability on the extraction and purification of artemisinin [6]. For this we also developed a mass spectrometry method for rapid evaluation of six key artemisinin-related metabolites in raw extracts [15]. Based on the latter work we were surprised to see a reproducibly high concentration of

* Corresponding author: Fax: +44 12234796

E-mail address: aal35@cam.ac.uk (A. Lapkin).

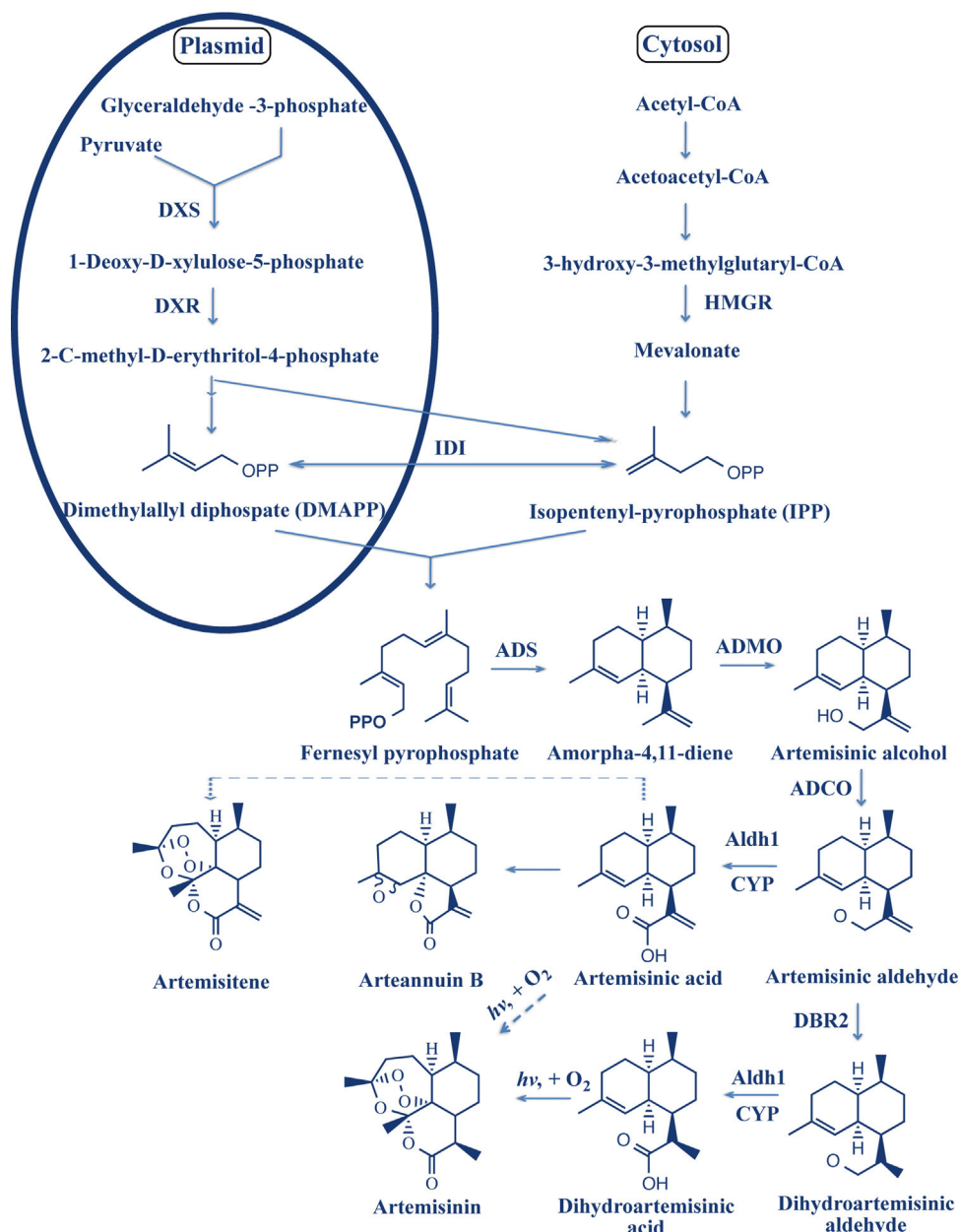
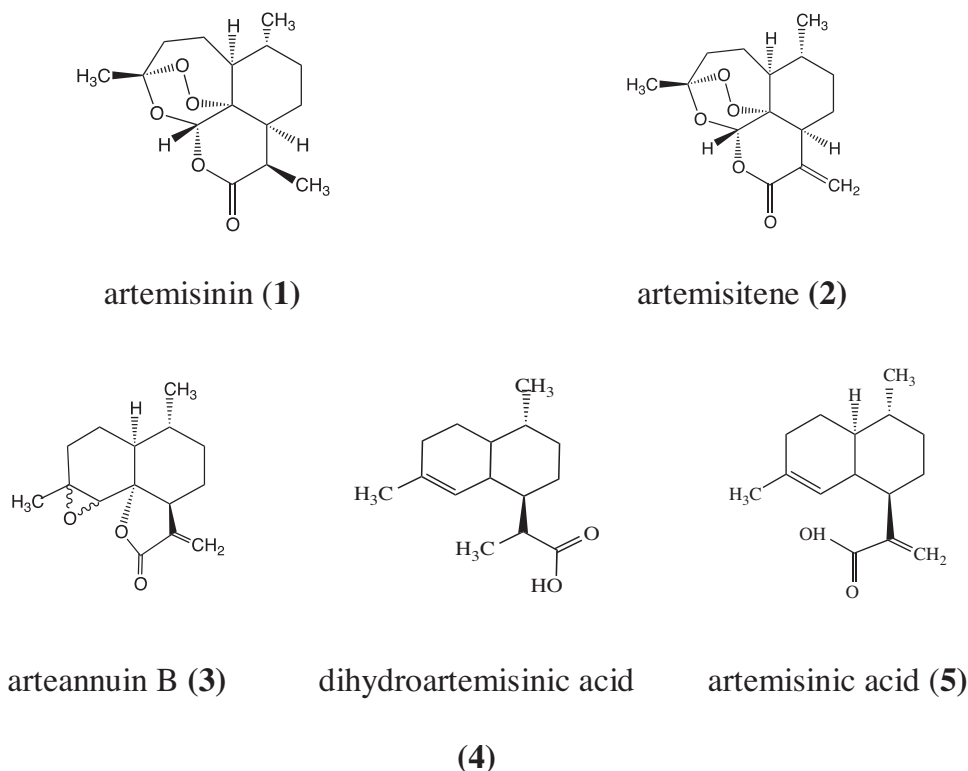


Fig. 1. Schematic of artemisinin biosynthetic pathway. DXS=1-deoxy-D-xylulose-5-phosphate synthase, DXR=1-deoxy-D-xylulose-5-phosphate reductoisomerase, IDI=IPP and DMAPP isomerase, HMGR=3-hydroxy-3-methylglutaryl-CoA reductase, ADS=Amorpha-4,11-diene synthase, ADMO=Amorpha-4,11-diene monooxygenase, ADCO=Amorpha-4,11-diene C-12 oxidase, DBR2=Double bond reductase 2, Aldh1=Aldehyde dehydrogenase 1, CYP=Cytochrome P450.

dihydroartemisinic acid, specifically in the biomasses profiled. This coincided with an extensive programme of testing of the impact of growth conditions, mainly light and temperature, on the production of artemisinin, undertaken in Madagascar. Based on the available literature data for the biosynthetic pathway to artemisinin, a high concentration of dihydroartemisinic acid may favour one of the two proposed routes [8]. The high concentration may also offer a new biomass exploitation route through synthetic conversion of dihydroartemisinic acid into artemisinin in a process similar to that developed by Seeberger [16]. Hence, we decided to undertake an extensive study of key artemisinin-related metabolites in the Madagascar-grown *A. annua* biomass as a function of growth stages and conditions. Attempts at studying the evolution of artemisinic metabolites in *A. annua* plants grown in glasshouses have been undertaken earlier [8,17,18]. This study distinctly assesses the profile of the *field*-grown cultivars

and evaluates possible biosynthetic associations among these metabolites.

To aid in the identification of associations between metabolites and biosynthetic precursors, different methods of multivariate data analysis have been used in the literature [8,19]. Numerous multivariate methods, both supervised and unsupervised, have been described in terms of reporting standards for metabolomics studies [20–22]. However, it was recently shown that some methods e.g., partial least squares-discriminant analysis (PLS-DA) [23], might lead to incorrect conclusions when used by inexperienced researchers [24,25]. Therefore, in this study we employed principal component analysis (PCA) [26], hierarchical cluster analysis (HCA) [27] and discriminant function analysis (DFA) [28], since these methods offer transparent outcomes. In addition, DFA in comparison to PLS-DA requires fewer decisions about parameters that have to be optimized and, therefore, the model is



Scheme 1. Chemical structures of metabolites of interest found in *A. annua*.

more directly related to the statistical properties of the data [24,25,29,30].

This paper reports the results of a detailed study of the evolution of key artemisinin-related metabolites for different *A. annua* cultivars grown in Madagascar. It reports an improved liquid chromatography-tandem mass spectrometry (LC-MS/MS) method for these key metabolites and presents results of detailed multivariate data analysis of the metabolite concentrations and growth conditions.

2. Experimental

2.1. Chemicals

Artemisinin reference standard (98%), LC-MS grade formic acid in water and acetonitrile, high-performance liquid chromatography (HPLC) grade acetonitrile, ethyl acetate and hexane were obtained from Sigma-Aldrich (Dorset UK). Dihydroartemisinic acid (>96%) was purchased from Apin Chemicals (Oxfordshire, UK). Arteannuin B, artemisitene and artemisinic acid were kindly provided by Walter Reed Army Institute of Research (Washington USA). Purified water ($\sim 18 \text{ M}\Omega \text{ cm}^{-1}$) was dispensed from a Milli Q system (Millipore, UK).

2.2. Analytical standards

Standard stock solutions of 1 mg mL^{-1} of artemisinin (1), artemisitene (2), arteannuin B (3), artemisinic acid (4) and dihydroartemisinic acid (5) in acetonitrile were prepared. The analytical standard was a mixture of all five standards in a mobile phase in the concentration range between $0.37\text{--}10 \mu\text{g mL}^{-1}$ for (1), (2) and (3). For (4) and (5) β -artemether was used as an internal standard at $5 \mu\text{g mL}^{-1}$.

2.3. Plant varieties

Thirteen varieties of *A. annua* plants from breeding programmes selecting for high artemisinin content were evaluated for artemisinin and related metabolites levels to provide crosses for new hybrids. These included four genotypes (CNAP1209R, CNAP5013, CNAP8001R, CNAP1252R) from the breeding programme of the Centre for Novel Agricultural Products (CNAP, York, UK). Another six (NIAB1125, NIAB1129, NIAB1105, NIAB1114, NIAB1118, NIAB1131) were from the National Institute of Agricultural Biology (NIAB, Cambridge, UK), and one cultivar (KEF1) from Kenya. These were benchmarked against the Apollon and NIAB1062 genotypes from Médiplant and NIAB respectively. *Laurus nobilis* (bay laurel) was used as a negative control and plant matrix for the assessment of the matrix effect.

2.4. Sampling period

Sampling for metabolic profiling commenced at 90 days post transplantation and samples were taken every fourth night up to 161 days. This window coincided with the period of quantifiable concentrations of all the targeted metabolites in the plant from earlier observations. Little or no flowering occurs in the plant during the winter season when these were grown, consequently corresponding to the vegetative and bolting stages of the plant.

2.5. Sample extraction and preparation

Extraction of dried plant leaves was carried out using published protocols [15,31,32]. Briefly, 0.5 g of the biomass was extracted with 10 mL of hexane modified with 5% (v/v) ethyl acetate in an ultrasonic bath (Kerry, Ultrasonics, UK), which was kept cold with ice; the extraction duration was always 30 min . The extracts were dried *in vacuo* and the residue re-suspended in acetonitrile before filtering through a $0.2 \mu\text{m}$ Nylon syringe filter (Fisher Scientific, UK). An

Table 1
TQD parameters for MS/MS experiments.

	Cone voltage (V)	Collision voltage (V)	MRM transitions
Artemisinin (1)	24	7	283 → 219 + 229 + 247 + 265
Artemisitene (2)	30	10	281 → 217 + 227 + 245 + 263
Arteannuin B (3)	28	9	249 → 189.5 + 213.5 + 221.5 + 231.5
Dihydroartemisinic acid (4)	32	12	237 → 191 + 201 + 219
Artemisinic acid (5)	32	12	235 → 190 + 200 + 218

aliquot of the filtrate was diluted in the mobile phase to an appropriate concentration and the internal standard added for LC–MS analysis. *L. nobilis* (bay laurel), which was used as a negative control, was extracted using the same procedure.

2.6. Plant metabolites analysis

2.6.1. Liquid chromatography method

The liquid chromatography analyses were performed using a Shimadzu Prominence HPLC system coupled to a Xevo tandem quadrupole mass spectrometer (Waters Corp., Milford, MA, USA). The Shimadzu system consisted of a binary pump, an auto-sampler, a UV detector and a column oven. The column heater was set at 30 °C and a Cortecs C18 column (100 mm × 2.1 mm, 2.7 μm) (Waters Corp., USA) was used for separation of the metabolites. The mobile phase consisted of A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. Chromatographic separation was achieved using a linear gradient: 0–7.0 min, 25–98% B; 7–9.5 min, 98% B; 9.5–10 min, 98–25% B; 10–15 min, 25% B; at a flow rate of 0.4 mL min⁻¹.

2.6.2. Multiple reaction monitoring (MRM) method

The tandem MS experiments were carried out in positive ionization mode (ESI+) and acquisition was performed in MRM mode. The cone and de-solvation gas flow rates were set at 45 and 800 Lh⁻¹, respectively, while the capillary voltage, the source and de-solvation temperatures were similar for all analytes at 40 kV, 150 and 350 °C respectively. The MS parameters were automatically defined using Waters IntelliStart® software for tuning and calibration of the tandem quadrupole detector (TQD) and subsequently was manually optimized. Final parameters are shown in Table 1. Quantification was achieved using MRM modes for the above transitions. The dwell time was automatically set at 0.161 s. Data were acquired by MassLynx v. 4.1 and processed for quantification with QuanLynx v. 4.1 (Waters Corp., Milford, MA, USA).

2.7. Multivariate data analysis

All statistical data analyses were conducted using the R (v. 3.1.0) software environment (<http://cran.r-project.org/>). Two unsupervised algorithms were used to reveal hidden structure in the data, namely PCA and HCA. These methods allow identification of relationships among different data points (i.e. samples) as well as between different variables, in this case metabolites [26,27]. However, both techniques only permit analysis of the data according to its features, i.e. without class labels and, therefore, DFA was employed to discover the patterns, which will link the inputs with the output [28].

2.8. Data pre-processing

In this study, samples were analysed using LC coupled to MS. However, this method provides data that are often incomplete, which may be related to a number of reasons, such as complex deconvolution, measurement error, detection sensitivity and

many more [33]. Therefore, pre-processing methods/strategies were employed to make the data more clear and suitable for analysis. In the first stage, all the missing values were replaced using random forests [33]. This method was selected over other approaches as this technique provides more robust and accurate results as recently shown [33]. Once all missing values were imputed, the data were auto-scaled to transform all the variables to one comparable scale [34].

3. Results and discussion

3.1. Chromatography

An earlier published method for the analysis of artemisinin, artemisitene artemisinic acid and arteannuin B was used as a starting point [15]. The level of dihydroartemisinic acid (DHAA) reported by this method is higher than that reported by other investigators [35–37]. This discrepancy informed a revisiting of our DHAA method [15] and investigation of why elevated readings for the metabolite were obtained earlier. Consequently we suspected that the transitions (237 → 190 + 200 + 218) chosen for the analysis of dihydroartemisinic acid in the earlier method [15] may be specific not only to dihydroartemisinic acid alone but may be susceptible to interference from very closely related compounds in the extract. Our attempt at the identification of these related metabolites is ongoing, while here we undertook to correct this error and develop a more robust method specific for dihydroartemisinic acid alone.

3.1.1. Method validation—dihydroartemisinic acid

The International Conference on Harmonization (ICH) [38] guideline for bio-analytical method validation were employed for the definition and determination of recovery specificity, precision, limits of detection and quantification and analytical range of the method.

3.1.2. Recovery

Eight equal portions (0.5 g) of one of the (CNAP) samples were used to assess the recovery of the metabolites from the extract. Three of the extracts were spiked with dihydroartemisinic acid (4) to give a final concentration of 4.10 μg mL⁻¹. The remaining five extracts were un-spiked. Table 2 shows that a recovery of between 97.56–108.05 % and a mean of 104.31% was obtained for the metabolite.

3.1.3. Specificity

Fig. 2A shows the fragmentation profile of dihydroartemisinic acid from a tandem MS scan. The MRM transitions selected for dihydroartemisinic acid are shown in Fig. 2B. Our non-selection of the comparatively stronger 237 > 163 *m/z* was informed by the reduced specificity associated with the lower *m/z* relative to higher ones. The intensities of the chosen transitions were adequate, see Fig. 2B and were used for the qualitative tandem (MS/MS) experiment to identify the metabolite. The total ion current (TIC) data was obtained from the sum of the *m/z* intensities of each of the transitions monitored while the transition with the relatively high-

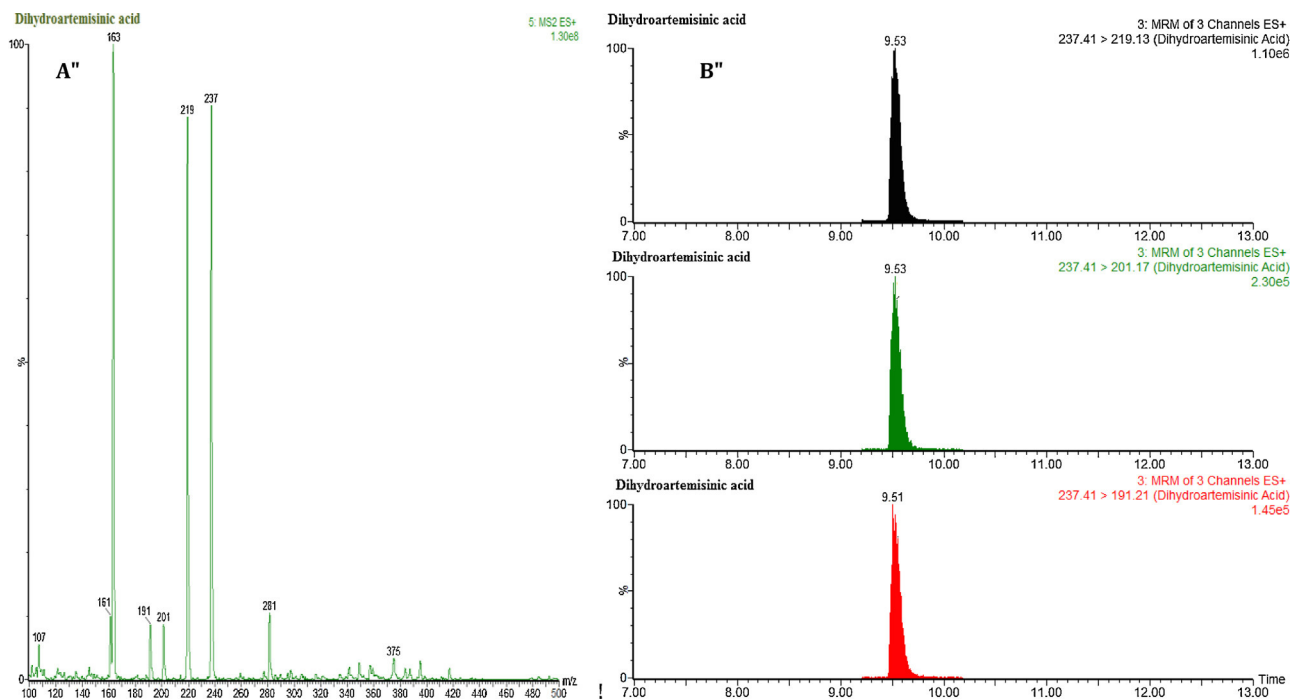


Fig. 2. (A) Fragmentation pattern for dihydroartemisinic acid. (B) Chosen MRM transitions for dihydroartemisinic acid.

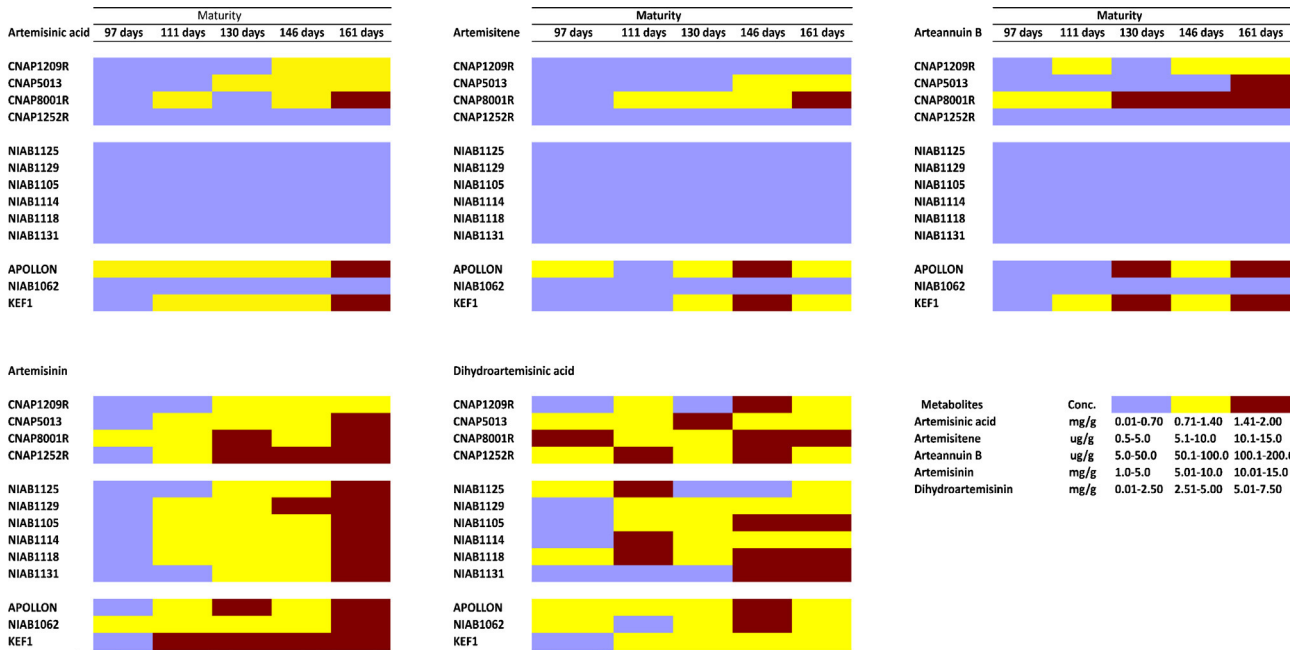


Fig. 3. A heat map of the artemisinin-related metabolites analysed in this study.

est signal intensity (237 > 219) of the chosen ions was used in the quantitative analysis. Specificity of MS/MS experiments is inherent; however further comparative analysis of two samples of equal concentration of DHAA (5 µg mL⁻¹), one sample spiked with closely related metabolites (artemisinin (1), artemisitene (2), arteannuin B (3) and artemisinic acid (5), each at 5 µg mL⁻¹) and the other un-spiked was carried out. Both samples showed only dihydroartemisinic acid peak in the MRM experiment for the metabolite and a mean concentration of 5.03 ± 0.03 µg mL⁻¹ was obtained.

3.1.4. Limit of detection (LOD), lower limit of quantification (LLOQ) and dynamic range The limit of detection is defined as the lowest amount of analyte in a sample, which can be detected but not necessarily quantitated, while the lower quantification limit is the lowest amount of analyte in a sample, which can be quantitatively determined with suitable precision and accuracy. The standard calibration curve was used to calculate these limits following Miller and Miller [39]. The LOD and LLOQ for the metabolite were determined to be 0.12 and 0.37 µg mL⁻¹ respectively, see Table 2.

Table 2

Recovery, LOD, LLOQ, precision and matrix effect.

Recovery				
	Mean quantity in un-spiked sample ^a ($\mu\text{g mL}^{-1}$)	Spiked quantity ^b ($\mu\text{g mL}^{-1}$)	Total quantity in spiked sample ^c ($\mu\text{g mL}^{-1}$)	Recovered quantities ($\mu\text{g mL}^{-1}$)
Spiked sample 1 ^d	0.60 \pm 0.06	4.10 \pm 0.05	4.60 \pm 0.17	4.00 \pm 0.12 (97.56%)
Spikes sample 2 ^d	0.87 \pm 0.07	4.10 \pm 0.05	5.30 \pm 0.10	4.43 \pm 0.20 (108.05%)
Spiked sample 3 ^d	1.47 \pm 0.12	4.10 \pm 0.05	5.87 \pm 0.43	4.40 \pm 0.55 (107.32%)
Mean + standard error				4.28 \pm 0.14 (104.31%)
LOD, LLOQ and precision				
	LOD ^e = 1.24×10^{-1} ($\mu\text{g mL}^{-1}$)	LLOQ ^e = 3.72×10^{-1} ($\mu\text{g mL}^{-1}$)	Regression equation $y = 59747x + 25816$	
Precision (%)		Mean	Range	Coefficient of variance (CV)
	Injection ^f	102.46	98.36–106.56 ($n = 10$)	4.22
	Within-day ^g	104.72	95.85–112.00 ($n = 12$)	8.02
	Between-day ^g	103.37	90.00–116.00 ($n = 48$)	8.80
Ion suppression (matrix effect)				
	Spiked plant matrix ($\mu\text{g mL}^{-1}$) ^h	Spiked blank matrix ($\mu\text{g mL}^{-1}$) ⁱ		
		Sample 1	Sample 2	Sample 3
Mean	3.32	3.54 (+2.79%)	3.38 (+0.33%)	3.45 (+1.07%)
Standard error (SE)	0.02	0.07	0.01	0.03

Eight equal samples of 0.5 g of dried *A. annua* leaves (CNAP variety) were extracted and prepared for analysis. ^aFive of these extracts were un-spiked while three were spiked at indicated levels^b. The total quantity^c of analyte in the samples was calculated as the sum of the mean quantities in five un-spiked samples and the spiked quantity. Analyte levels in individual spiked samples were determined^d and absolute and percentage recoveries presented (in bracket). ^eLOD and LLOQ calculations were based on 10 point calibration graph and the following formulae $\text{LOD} = Y_B + 3S_B$ and $\text{LLQD} = 3\text{LOD}$, where Y_B is the blank signal (the y -intercept) and the S_B is the standard deviation of the blank (the random error in the y -direction) [39]. ^fInjection precision was assessed by n determinations at 100% concentration. ^gWithin and between-day precision were determined over 5 concentration levels covering the calibration range for both precisions. ^hMean of three determinations of spiked standards at 3.32 mg mL⁻¹. ⁱThree determinations of blank matrix (mobile phase; 0.1% formic; 0.1% formic in acetonitrile, 75:25) spiked with standard at an equivalent concentration to spiked plant matrix. Percentage enhancement is shown in brackets.

The linearity of the calibration curve was above $r^2 > 0.99$ in both the mobile phase and the mobile phase spiked with the matrix. The regression and sensitivity indices presented in Table 2 were obtained for the standards prepared in the mobile phase spiked with plant matrix. The method's dynamic range was 0.12–10 $\mu\text{g mL}^{-1}$.

3.1.5. Precision

The precision parameters for the method are shown in Table 2. The injection precision was assessed from 10 determinations of the metabolite at 100% of the concentration in the analytical sample in a single day. The coefficient of variation (CV) for these determinations was found to be adequate (4.2%).

Within day precision was determined for 6 concentration levels covering the analyte calibration range, making a total of 12 analyses on a single day. Between days accuracy was calculated for the same calibration range spread over 3 months, resulting in a total of 48 determinations. An accuracy range of 90.00–116.00% was obtained for both within and between day determinations and the coefficient of variance was less than 9.0% in both analyses.

3.1.6. Ion suppression and matrix effect

The suppression or enhancement of ionization due to the presence of endogenous components in the plant matrix may occur in LC–MS or LC–MS/MS based assays with consequential impact on the precision and accuracy of the bio-analytical method [40]. Different strategies have been used for the assessment of the matrix effect [41,42], including the spike method that was used in our determination. The strategy involves comparing the response of the metabolite in the plant matrix to the response in the blank matrix (mobile phase), both matrices spiked with an equivalent concentration of the metabolite and treated through the sample preparation protocol [43].

The three samples of dihydroartemisinic acid spiked with plant matrix Table 2 showed minimal ion enhancement (0.33–2.79%)

compared with the metabolites solubilised in the blank matrix (mobile phase).

The above validation indices showed the method to be robust and sensitive for the analysis of dihydroartemisinic acid in crude extracts. The levels of the metabolite detected by this method is consistent with that reported by Larson et al., citing a higher artemisinin content relative to dihydroartemisinic acid for most of the commercially grown lines they evaluated [17]. This is in contrast to the previous protocol [15] that found higher concentrations of dihydroartemisinic acid relative to artemisinin in the analysed biomasses. Accumulation of secondary metabolites in plants depends on various factors including, among others, stage and condition of growth. The assessment of the evolutionary trend for some artemisinic metabolites in *A. annua* by both Wang et al. [8] and Towler and Weathers [18] showed a lower accumulation of dihydroartemisinic acid relative to artemisinin at the later vegetative stage of the plant, this stage, coincided with our sampling. They also observed, that the trend was reversed at the early vegetative stages.

3.2. Heat map and metabolite profile

The metabolic profile of artemisinin related compounds over the growing period of the experimental plants is visualized as a heat-map in Fig. 3. The map is colour-coded to three levels of metabolite expression (blue = low range, yellow = middle range, brown = high range). The categorisation is based on the determined range of the metabolite content found in all the samples analysed.

The NIAB samples showed very close similarity in their artemisinic acid, artemisitene and arteannuin B content. The levels of these metabolites do not change significantly over the growth period compared to other cultivars sampled, and are all detected at the comparatively low concentration range for these metabolites. The CNAP samples showed comparatively more variation among cultivars with significant increases in the content of these metabolites detected over the sampled growth period in

Table 3
Content of artemisinin and related metabolites at the end of sampling.

	Plant cultivar	Artemisinin/mg g ⁻¹	Dihydroartemisinic acid/mg g ⁻¹	Artemisinic acid/mg g ⁻¹	Artemisitene/μg g ⁻¹	Arteannuin B/μg g ⁻¹
1	V1-CNAP1209R	8.47 ± 0.51	4.95 ± 0.60	0.71 ± 0.04	2.63 ± 0.13	65.23 ± 0.52
2	V2-CNAP5013	11.48 ± 0.28	4.98 ± 0.51	1.17 ± 0.07	9.80 ± 1.37	103.62 ± 4.50
3	V3-CNAP8001R	13.01 ± 0.08	5.01 ± 0.54	1.46 ± 0.08	13.47 ± 0.63	146.83 ± 11.25
4	V4-CNAP1252R	12.71 ± 0.44	4.92 ± 0.61	0.45 ± 0.05	4.94 ± 0.57	40.23 ± 4.12
5	V5-KF1	12.21 ± 0.88	4.86 ± 0.40	1.65 ± 0.13	6.84 ± 0.44	159.97 ± 10.47
6	V6-NIAB1125	10.80 ± 0.26	4.74 ± 0.49	0.53 ± 0.03	4.31 ± 0.46	34.45 ± 2.03
7	V7-NIAB1129	11.10 ± 0.30	3.73 ± 0.49	0.41 ± -0.03	3.93 ± 0.69	37.22 ± 4.03
8	V8-NIAB1105	11.07 ± 0.26	5.35 ± 0.41	0.58 ± 0.02	2.45 ± 0.26	49.67 ± 1.40
9	V9-NIAB1114	10.04 ± 0.51	4.05 ± 0.69	0.42 ± 0.02	2.84 ± 0.12	45.21 ± 1.45
10	V10-NIAB1118	11.27 ± 0.12	7.36 ± 0.11	0.60 ± 0.02	3.87 ± 0.04	32.47 ± 4.21
11	V11-NIAB1131	10.40 ± 0.65	5.08 ± 0.38	0.39 ± 0.01	3.08 ± 0.35	33.80 ± 3.95
12	T1-APOLLON	10.91 ± 0.15	4.67 ± 0.28	1.81 ± 0.06	10.97 ± 0.79	171.32 ± 7.45
13	T2-NIAB1062	10.48 ± 0.28	4.80 ± 0.48	0.58 ± 0.04	2.63 ± 0.45	48.35 ± 2.55

some of the cultivars. This is also true for the Apollon and Kenyan cultivars sampled, see Fig. 3.

The artemisinin and dihydroartemisinic acid content of the sampled cultivars increased over the growth period in all of the varieties with the exception of one CNAP cultivar (CNAP8001R), which had high levels of DHAA at the beginning and end of sampling.

The heat maps for artemisinic acid, artemisitene and arteannuin B are generally similar for these metabolites in the analysed samples. The artemisinin and dihydroartemisinic acid maps are more distinct from the other metabolites and these distinctions seem to correlate with the result of the principal component analysis below.

3.3. Metabolic profile and biosynthetic association

Table 3 shows the metabolic profiles of sampled cultivars at the end of the sampling period. For this period, the highest accumulation of artemisinin ($13.01 \pm 0.08 \text{ mg g}^{-1}$), dihydroartemisinic acid ($7.36 \pm 0.11 \text{ mg g}^{-1}$), artemisinic acid ($1.81 \pm 0.06 \text{ mg g}^{-1}$), artemisitene ($13.47 \pm 0.63 \text{ μg g}^{-1}$) and arteannuin B ($171.32 \pm 7.45 \text{ μg g}^{-1}$), occurred in CNAP8001, NIAB1118, Apollon, CNAP8001 and NIAB1118, respectively. Overall CNAP8001 had the best profile for the metabolites analysed.

The comparatively higher level of accumulation of dihydroartemisinic acid relative to artemisinic acid in these cultivars is consistent with the observation by Wallaart et al., who showed that a higher accumulation of dihydroartemisinic acid was accompanied by higher artemisinin content, while a negative correlation existed between accumulation of artemisinic acid and artemisinin [14].

In the 13 cultivars examined, artemisinin is accumulated at a level of between 2 and 3 (2.4 ± 0.08) times the levels of dihydroartemisinic acid, except in two cultivars (NIAB1118=1.5 and CNAP1209=1.7 times). Interestingly, a similar consistency was observed in the relative accumulation of artemisinic acid and arteannuin B, which was between 10–12 (10.9 ± 0.20) times (except in NIAB1118=18 and NIAB1125=15 times). This consistency was not observed in any other coupling and may further suggest an association between artemisinin and dihydroartemisinic acid in the biosynthetic pathways of our sampled genotypes, while artemisinic acid seems to be associated with arteannuin B in these cultivars.

Some investigators [44,45], have proposed arteannuin B as the intermediate in the pathway from artemisinic acid to artemisinin. However, Brown and Sy in their *in vivo* feeding experiments with artemisinic acid labeled with both heavy carbon (¹³C) and proton (²H), excluded it as precursor in biosynthesis of artemisinin in *A. annua*, but showed dihydroartemisinic acid as the late stage precursor of artemisinin [46].

3.4. Principal component analysis

A total of 388 metabolite profiles obtained from 13 *A. annua* cultivars were evaluated by PCA. These sets of data were divided into five distinct groups, corresponding to the stages of growth sampled. Each group comprises an average of 77 samples. Fig. 4A is the scores plot of samples which show metabolite clustering based on stages of growth. The relationships between monitored metabolites are shown in Fig. 4B. Close association is observed for artemisinic acid and arteannuin B with artemisitene the next associated to these two.

3.5. Hierarchical cluster analysis

HCA was conducted using function “hclust” in R’s “stats” package in order to indicate relationship between the metabolites investigated and to confirm our findings relating to the correlated variables as investigated using PCA. This is a frequently used method that allows identification of similarities/dissimilarities between different individuals or correlations between different variables, here—metabolites. The algorithm calculates a distance matrix for a given data set using the “Euclidean distance”. As long as the distance between any observations is calculated, the “Ward” linkage is used to link different elements into clusters [27]. The output of the hierarchical clustering computations is represented as a tree-structured graph called a dendrogram. The hierarchical structure can either depict the assemblage for observations according to their characteristics or display correlations between features, *i.e.* metabolites, as calculated using correlation coefficients. Consequently, the variables with a short in-between distance are expected to be similar or related.

The results of HCA for the metabolites are shown in Fig. 5, where the dendrogram can be divided into two subgroups: the first group contains arteannuin B which is highly correlated with artemisinic acid followed by artemisitene and in the second group, dihydroartemisinic acid and artemisinin, shows a much smaller correlation in comparison to the first group. The results of the cluster analysis confirm what has been observed when looking at the loadings plot, Fig. 4B, *i.e.* that two subgroups of related metabolites can be detected.

3.6. Discriminant function analysis

PCA is not always capable of satisfactorily separating the variations produced by each factor. This can be related to the fact that the variations caused by these factors are spread across a different number of components, which makes the results somewhat problematic to read. Therefore, in Fig. 4A some overlapping could be observed. In order to circumvent this scenario supervised methods can be applied to make the results clearer. One of the frequently

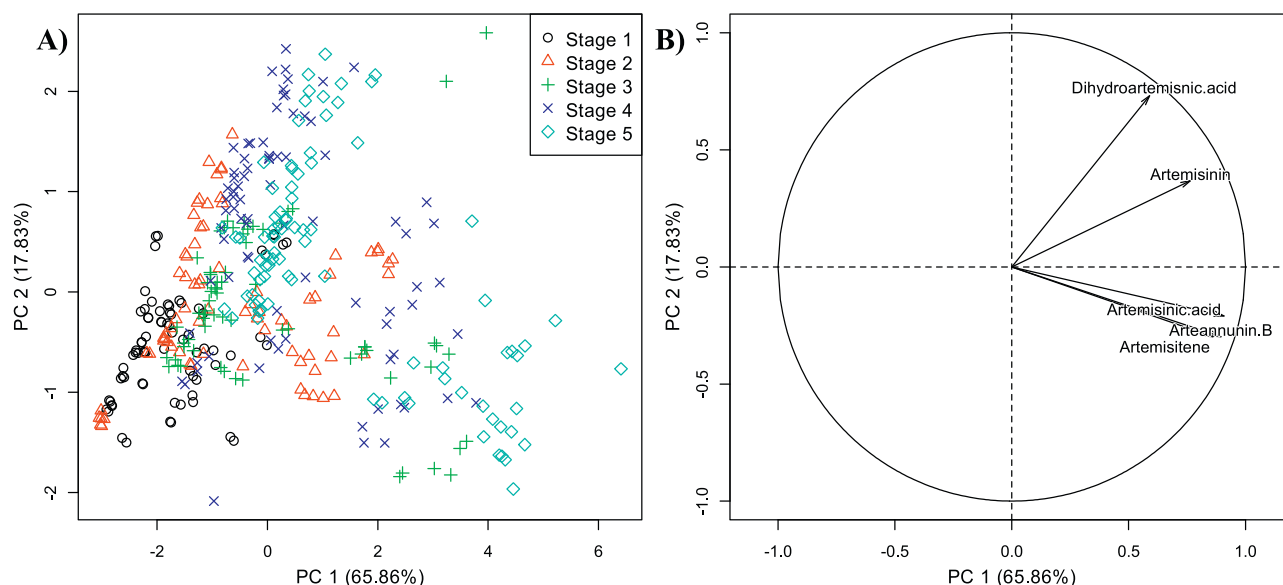


Fig. 4. Results of PCA. (A) The scores plot showing samples clustered according to different stages of growth. (B) A loadings plot displaying relationships between different variables, in this case the metabolites of interest.

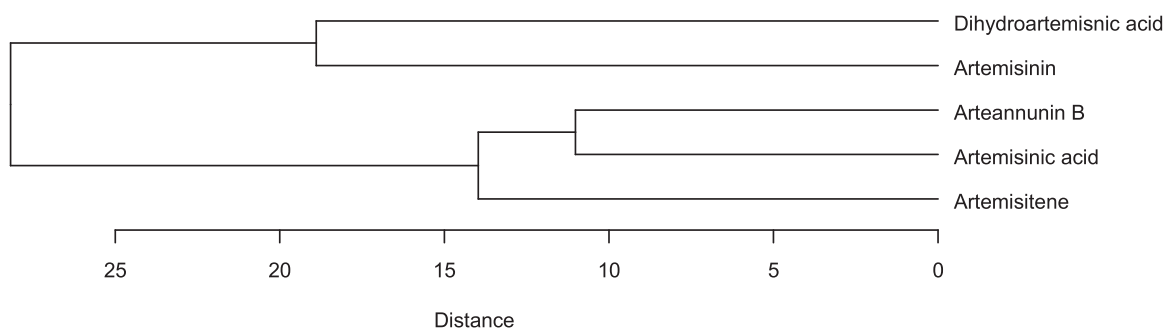


Fig. 5. A dendrogram based on "Ward" linkage representing five different metabolites.

used methods to reveal the structure in data with known labeling is DFA. This technique applies Fisher's ratio so as to search for a linear combination of factors that minimizes within group variance and maximizes between group variance [28]. This allows for allocation of objects into groups, such that objects in the same group are related.

The scores plot of DFA, see Fig. 6, indicates that the algorithm was able to clearly separate samples related to different stages of plant growth. In addition, the grouping is mostly related to the levels of artemisinin, the metabolite that was recognized as an influential factor using both unsupervised approaches (PCA and HCA). The level of artemisinin strongly depends on the growth stage, with the highest levels observed in the last stage.

Overall, the combined use of these advanced multivariate tools to study the evolution of artemisinic metabolites showed two distinct sub-groupings of the metabolites. Artemisinin and dihydroartemisinic were clearly associated, and there was a correlation for artemisinic acid, arteannuin B and artemisitene. The close proximity of artemisinic acid and arteannuin B, and to a lesser extent for artemisinin and dihydroartemisinic acid on the dendrogram along with the consistency of their relative accumulation in the analysed samples Table 3, supports the direct link between these metabolites in the biosynthetic pathway Fig. 1. Higher accumulation of artemisinin has also been associated with lower concentrations of arteannuin B and artemisinic acid and *vice versa* [8,47]. Our findings

strengthen the suggestion of different chemotypes in *A. annua*: at least one of high artemisinin and low artemisinic acid content and the other of low artemisinin and high artemisinic acid levels [8,47]. Consequently, with carbon flow directed at these two alternative or competing paths in the later stage of the biosynthetic process, a key to the favoured direction of carbon flow may be the double bond reductase 2 (DBR2) enzyme which catalyses the reduction of artemisinic aldehyde to dihydroartemisinic aldehyde for subsequent conversion to dihydroartemisinic acid. Targeted study of the enzyme in the different chemo-types could be advantageous.

4. Conclusions

In this study 13 genotypes were screened from three breeding programs for superior artemisinin related metabolic profiles using an earlier published and updated LC–MS protocols. Generally, all evaluated validation indices showed the new analytical method to be robust and sensitive for the analysis of the metabolites in crude extracts.

The data set obtained from the metabolic profiling of the genotypes was subjected to multivariate analysis to identify association and correlation among metabolites. The PCA loadings plot and the dendrogram shows two distinct grouping among the five metabolites analysed. Artemisinic acid and arteannuin B were found to be more closely related than any of the metabolites, with artemisitene

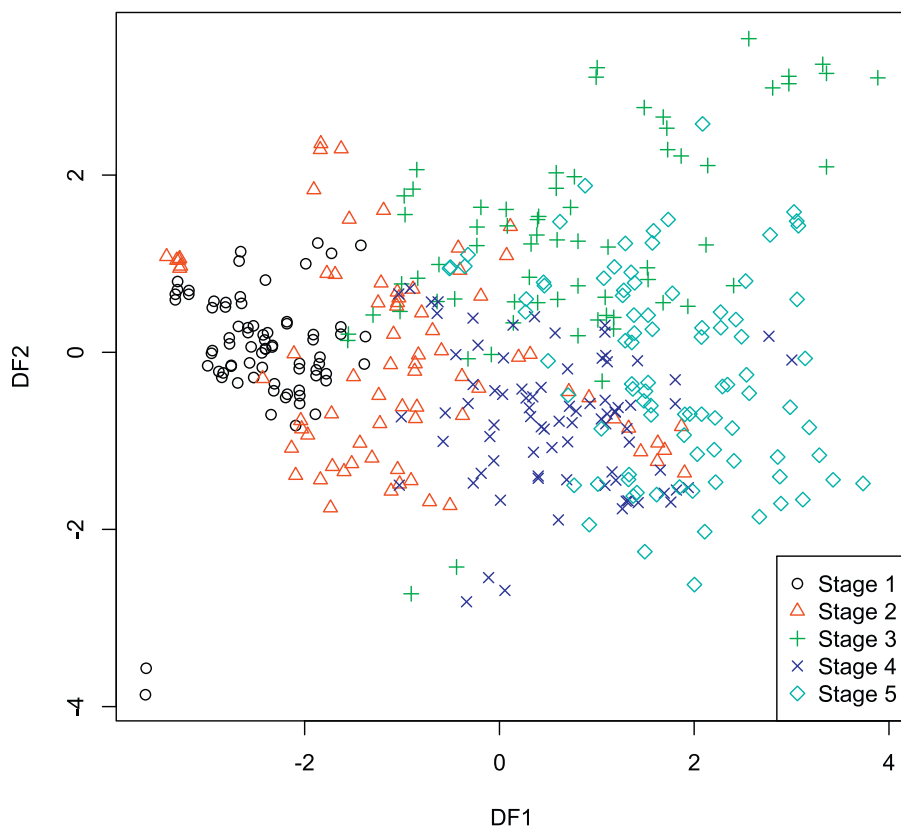


Fig. 6. DFA scores plot, showing samples clustered according to different stages of plant growth.

correlating to these two rather than with artemisinin and dihydroartemisinic acid. Among the metabolites, artemisinin was the closest to dihydroartemisinic acid. A visualized heat map and a rather consistent ratio of accumulation of these coupled metabolites tend to add to the validity of the PCA associations.

Acknowledgements

The research leading to these results has received funding from Engineering and Physical Sciences Research Council project “Closed Loop Optimization for Sustainable Chemical Manufacture” [EP/L003309/1]. We are grateful to Medicines for Malaria Ventures for providing funding for HPLC instrument and University of Cambridge for co-funding the Xevo TQD instrument. Plant biomass samples were kindly provided by Bionexx Ltd, Madagascar.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jpba.2015.10.003>.

References

- [1] WHO, World Malaria Report 2014, Geneva: World Health Organisation; 2014. http://www.who.int/malaria/publications/world_malaria_report_2014/report/en/, (2014) 242.
- [2] WHO, Who expert committee on specifications for pharmaceutical preparations, World Health Organisation Geneva Switzerland, 44th Report (2010) 276.
- [3] WHO, Quality requirements for artemisinin as a starting material in the production of antimalarial active pharmaceutical ingredients (APIs)—Revised Draft for Comment, World Health Organisation Geneva Switzerland, Working document QAS/10.349 (2011).
- [4] WHO, World Malaria Report 2012, Geneva: World Health Organisation; 2012. http://www.who.int/malaria/publications/world_malaria_report_2012/report/en/index.html, (2012) 1–195.
- [5] J. Turconi, F. Griollet, R. Guevel, G. Oddon, R. Villa, A. Geatti, M. Hvala, K. Rossen, R. Göller, A. Burgard, Semisynthetic artemisinin, the chemical path to industrial production, *Org. Process Res. Dev.* 18 (2014) 417–422.
- [6] A. Lapkin, E. Adou, B.N. Mlambo, S. Chemat, J. Suberu, A.E. Collis, A. Clark, G. Barker, Integrating medicinal plants extraction into a high-value biorefinery: an example of *Artemisia annua* L., *C. R. Chim.* 17 (2014) 232–241.
- [7] G.D. Brown, L.-K. Sy, Synthesis of labelled dihydroartemisinic acid, *Tetrahedron* 60 (2004) 1125–1138.
- [8] H. Wang, C. Ma, L. Ma, Z. Du, H. Ye, G. Li, B. Liu, G. Xu, Secondary metabolic profiling and artemisinin biosynthesis of two genotypes of *Artemisia annua*, *Planta Med.* 75 (2009) 1625–1633.
- [9] G.D. Brown, The biosynthesis of artemisinin (Qinghaosu) and the phytochemistry of *Artemisia annua* L. (Qinghao), *Molecules* 15 (2010) 7603–7698.
- [10] J. Kirby, J.D. Keasling, Biosynthesis of plant isoprenoids: perspectives for microbial engineering, *Ann. Rev. Plant Biol.* 60 (2009) 335–355.
- [11] Y. Li, H. Huang, Y.-L. Wu, Qinghaosu (Artemisinin)—a fantastic antimalarial drug from a traditional Chinese medicine, in: *Medicinal Chemistry of Bioactive Natural Products*, John Wiley & Sons, Inc., 2006, pp. 183–256.
- [12] P.S. Covello, K.H. Teoh, D.R. Polichuk, D.W. Reed, G. Nowak, Functional genomics and the biosynthesis of artemisinin, *Phytochemistry* 68 (2007) 1864–1871.
- [13] R.K. Haynes, From artemisinin to new artemisinin antimalarials: biosynthesis, extraction, old and new derivatives, stereochemistry and medicinal chemistry requirements, *Curr. Top. Med. Chem.* 6 (2006) 509–537.
- [14] T.E. Wallaart, N. Pras, A.R.C. Beekman, W.J. Quax, Seasonal variation of artemisinin and its biosynthetic precursors in plants of *Artemisia annua* of different geographical origin: proof for the existence of chemotypes, *Planta Med.* 66 (2000) 57–62.
- [15] J. Suberu, L. Song, S. Slade, N. Sullivan, G. Barker, A.A. Lapkin, A rapid method for the determination of artemisinin and its biosynthetic precursors in *Artemisia annua* L. crude extracts, *J. Pharm. Biomed. Anal.* 84 (2013) 269–277.
- [16] F. Levesque, P.H. Seeberger, Continuous-flow synthesis of the anti-malaria drug artemisinin, *Angew. Chem. Int. Ed.* 51 (2012) 1706–1709.
- [17] T.R. Larson, C. Branigan, D. Harvey, T. Penfield, D. Bowles, I.A. Graham, A survey of artemisinic and dihydroartemisinic acid contents in glasshouse and global field-grown populations of the artemisinin-producing plant *Artemisia annua* L., *Ind. Crops Prod.* 45 (2013) 1–6.
- [18] M.J. Towler, P.J. Weathers, Variations in key artemisinic and other metabolites throughout plant development in *Artemisia annua* L. for potential therapeutic use, *Ind. Crops Prod.* 67 (2015) 185–191.
- [19] B.K. Neoh, H.F. Teh, T.L.M. Ng, S.H. Tiong, Y.M. Thang, M.A. Ersad, M. Mohamed, F.T. Chew, H. Kulaveerasingam, D.R. Appleton, Profiling of

- metabolites in oil palm mesocarp at different stages of oil biosynthesis, *J. Agric. Food Chem.* 61 (2013) 1920–1927.
- [20] P.S. Gromski, Y. Xu, E. Correa, D.I. Ellis, M.L. Turner, R. Goodacre, A comparative investigation of modern feature selection and classification approaches for the analysis of mass spectrometry data, *Anal. Chim. Acta* 829 (2014) 1–8.
- [21] M.M.W.B. Hendriks, F.A. van Eeuwijk, R.H. Jellema, J.A. Westerhuis, T.H. Reijmers, H.C.J. Hoefsloot, A.K. Smilde, Data-processing strategies for metabolomics studies, *Trac-Trends Anal. Chem.* 30 (2011) 1685–1698.
- [22] K.H. Liland, Multivariate methods in metabolomics—from pre-processing to dimension reduction and statistical analysis, *Trac-Trends Anal. Chem.* 30 (2011) 827–841.
- [23] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130.
- [24] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, *J. Chemom.* 28 (2014) 213–225.
- [25] P.S. Gromski, H. Muhamadali, D.I. Ellis, Y. Xu, E. Correa, M.L. Turner, R. Goodacre, Metabolomics and partial least squares-discriminant analysis: a marriage of convenience or a shotgun wedding?, *Anal. Chim. Acta*, (2015). in print.
- [26] H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.* 24 (1933) 417–441.
- [27] B. Everitt, Cluster Analysis, Heinemann Educational Books, London, UK, 1974.
- [28] B.F.J. Manly, Multivariate Statistical Methods: A Primer, Chapman and Hall, Boca Raton, 1986.
- [29] P.S. Gromski, E. Correa, A.A. Vaughan, D.C. Wedge, M.L. Turner, R. Goodacre, A comparison of different chemometrics approaches for the robust classification of electronic nose data, *Anal. Bioanal. Chem.* 406 (2014) 7581–7590.
- [30] J.A. Westerhuis, H.C.J. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J.J. van Velzen, J.P.M. van Duijnhoven, F.A. van Dorsten, Assessment of PLS-DA cross validation, *Metabolomics* 4 (2008) 81–89.
- [31] A.A. Lapkin, A. Walker, N. Sullivan, B. Khambay, B. Mlambo, S. Chemat, Development of HPLC analytical protocols for quantification of artemisinin in biomass and extracts, *J. Pharm. Biomed. Anal.* 49 (2009) 908–915.
- [32] C. A. Mannan, P.R. Liu, M.J. Arsenault, D.R. Towler, A. Vail, P.J. Weathers Lorence, DMSO triggers the generation of ROS leading to an increase in artemisinin and dihydroartemisinin acid in *Artemisia annua* shoot cultures, *Plant Cell Rep.* 29 (2010) 143–152.
- [33] P.S. Gromski, Y. Xu, H.L. Kotze, E. Correa, D.I. Ellis, E.G. Armitage, M.L. Turner, R. Goodacre, Influence of missing values substitutes on multivariate analysis of metabolomics data, *Metabolites* 4 (2014) 433–452.
- [34] P.S. Gromski, Y. Xu, K.A. Hollywood, M.L. Turner, R. Goodacre, The influence of scaling metabolomics data on model classification accuracy, *Metabolomics* 11 (3) (2015) 684–695.
- [35] J.F.S. Ferreira, Nutrient deficiency in the production of artemisinin, dihydroartemisinin acid, and artemisinin acid in *Artemisia annua* L, *J. Agric. Food Chem.* 55 (2007) 1686–1694.
- [36] J.F. Ferreira, J.M. Gonzalez, Analysis of underivatized artemisinin and related sesquiterpene lactones by high-performance liquid chromatography with ultraviolet detection, *Phytochem. Anal.* 20 (2009) 91–97.
- [37] B. Avula, Y.-H. Wang, T.J. Smillie, W. Mabusela, L. Vincent, F. Weitz, I.A. Khan, Comparison of LC–UV, LC–ELSD and LC–MS methods for the determination of sesquiterpenoids in various species of artemisia, *Chromatographia* 70 (2009) 797–803.
- [38] G.I.C.H.T., Validation of analytical procedures: text and methodology Q2 (R1), IUPAC: Geneva, (2005).
- [39] J.N. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, Prentice Hall, 2005.
- [40] C. Côté, A. Bergeron, J.-N. Mess, M. Furtado, F. Garofolo, Matrix effect elimination during LC–MS/MS bioanalytical method development, *Bioanalysis* 1 (2009) 1243–1257.
- [41] R. Bonfiglio, R.C. King, T.V. Olah, K. Merkle, The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds, *Rapid Commun. Mass Spectrom.* 13 (1999) 1175–1185.
- [42] B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC–MS/MS, *Anal. Chem.* 75 (2003) 3019–3030.
- [43] A. Van Eeckhaut, K. Lanckmans, S. Sarre, I. Smolders, Y. Michotte, Validation of bioanalytical LC–MS/MS assays: evaluation of matrix effects, *J. Chromatogr. B, Anal. Technol. Biomed. Life Sci.* 877 (2009) 2198.
- [44] R.J. Roth, N. Acton, A simple conversion of artemisinin acid into artemisinin, *J. Natl. Prod.* 52 (1989) 1183–1185.
- [45] M. Nair, D. Basile, Bioconversion of arteannuin B to artemisinin, *J. Natl. Prod.* 56 (1993) 1559–1566.
- [46] G.D. Brown, L.-K. Sy, In vivo transformations of artemisinin acid in *Artemisia annua* plants, *Tetrahedron* 63 (2007) 9548–9566.
- [47] T.E. Wallaart, N. Pras, A.C. Beekman, W.J. Quax, Seasonal variation of artemisinin and its biosynthetic precursors in plants of *Artemisia annua* of different geographical origin: proof for the existence of chemotypes, *Planta Med.* 66 (2000) 57–62.